Information Sciences 184 (2012) 155–165

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

On the properties of concept classes induced by multivalued Bayesian networks ☆

Youlong Yang *, Yan Wu

Department of Mathematics, School of Science, Xidian University, Xi'an 710071, PR China

ARTICLE INFO

Article history:
Received 11 October 2009
Received in revised form 15 August 2011
Accepted 21 August 2011
Available online 6 September 2011

Keywords:
Bayesian networks
Classification
Multivalued Bayesian networks
Concept classes
VC dimension
Inner product space

ABSTRACT

The concept class $\mathcal{C}_{\mathcal{N}}$ induced by a Bayesian network $\mathcal{N}$ can be embedded into some Euclidean inner product space. The *Vapnik–Chervonenkis* (VC)-dimension of the concept class and the minimum dimension of the inner product space are very important indicators for evaluating the classification capability of the Bayesian network. In this paper, we investigate the properties of the concept class $\mathcal{C}_{\mathcal{N}^k}$ induced by a multivalued Bayesian network $\mathcal{N}^k$, where each node X_i of $\mathcal{N}^k$ is a k-valued variable. We focus on the values of two dimensions: (i) the *VC*-dimension of the concept class $\mathcal{C}_{\mathcal{N}^k}$, denoted as $VCdim(\mathcal{N}^k)$, and (ii) the minimum dimension of the inner product space into which $\mathcal{C}_{\mathcal{N}^k}$ can be embedded. We show that the values of these two dimensions are $k^n - 1$ for fully connected k-valued Bayesian networks $\mathcal{N}_F^k$ with n variables. For non-fully connected k-valued Bayesian networks $\mathcal{N}^k$ without *V*-structure, we prove that the two dimensional values are $(k-1)\sum_{i=1}^{n} k^{m_i} + 1$, where m_i denotes the number of parents for the i^{th} variable. We also derive the upper and lower bounds on the minimum dimension of the inner product space induced by non-fully connected Bayesian networks.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Bayesian networks (BNs), also referred to as Belief Networks or Causal Networks, are directed graphical models for representing the probabilistic relationships between variables. The network structure is a directed acyclic graph (DAG) where each node represents a random variable [18,8,16,10]. BNs are a powerful tool for modeling the decision-making process under uncertainty. They have been successfully applied to various fields including machine learning and bioinformatics, and found particularly useful in knowledge representation, reasoning and learning under uncertainty [12,9,5].

Kitakoshi et al. [14] presented a reinforcement learning system that adapts to environmental changes using a mixture of BNs. Yang et al. [27] proposed a driver fatigue recognition model based on the information fusion technique and a dynamic BN. BNs are also widely used for classification due to their simplicity and accuracy [4,7,11,17]. Some approaches combine kernel methods and probabilistic models, such as Ben et al. [3], Taskar et al. [21], Gurwicz and Lerner [13], Chechik et al. [6], Aritz et al. [2] and Theodoros and Mark [22]. Altun et al. [1] proposed a kernel for the Hidden Markov Model, which is a special case of a BN. These methods study not only the domain-specific design of some probabilistic models, but also the information extracted from the data during the training of the probabilistic models.

For a data mining algorithm, the generalization capability is a very important feature that is often used to evaluate the performance of the algorithm. To improve the generalization of data mining algorithms, Wang and Dong [23] and Wang

☆ This work was supported by National Natural Science Foundation of China (61075055).

* Corresponding author.
E-mail addresses: youlongy@gmail.com, ylyang@mail.xidian.edu.cn (Y. Yang), yanwu@mail.xidian.edu.cn (Y. Wu).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.08.031

et al. [24] proposed a new refinement method based on a fuzzy decision tree, the rough set technique, and the fuzzy entropy. Based on the analysis for multi-valued attribute and multi-labeled data, Yi et al. [28] presented a new decision tree classification algorithm. In fact, decision tree is a special type of graphical model. To improve the power of the nearest neighbor-based algorithms to high dimensional data, a local learning method was proposed in the paper Tang et al. [20] for image processing. As machine learning techniques become more widely adopted regarding the classification and the structure prediction, it becomes increasingly important how to balance the computational consumption and the classification capability [19]. For the evaluation of the classification capability induced by a BN without considering the training data and the classification algorithm, two important indexes are often considered, i.e. *VC*-dimension and the smallest dimension of the inner product space (see Fig. 1). For example, to classify toys, a general procedure is to build a BN using the attributes of the toys or an expert system, and then train the data and design an appropriate algorithm. In this paper, we focus on the classification capability induced by a BN without considering the training data, the algorithm and the construction of the network graph. We will discuss the two dimensional-values by multivalued BNs.

For a BN $\mathcal{N}$, one can get a concept class $\mathcal{C}_{\mathcal{N}}$ induced by it. The concept class can be embedded into some Euclidean inner product space. Therefore, there are two dimensional-values: the *VC*-dimension, denoted as $VCdim(\mathcal{N})$ of the concept class $\mathcal{C}_{\mathcal{N}}$, and the minimum dimension of the inner product space, denoted as $Edim(\mathcal{N})$, into which $\mathcal{C}_{\mathcal{N}}$ can be embedded, as shown in Fig. 1. These are very important indexes to assess the classification capability of the BN. An interesting question has been raised as to whether these two dimensional-values are equal for BNs.

A network graph is fully connected if there is a directed edge between any two nodes, as illustrated in Fig. 2. Let X, Z, Y be three nodes of a network graph $\mathcal{G} = (V, E)$. We call (X,Z,Y) a *V*-structure if (1) $\mathcal{G}$ contains the edges $X \rightarrow Z$ and $Z \leftarrow Y$, and (2) X and Y are not adjacent in $\mathcal{G}$ [8]. For example, (X_3,X_5,X_4) in Fig. 3(a), (X_1,X_3,X_2) in Fig. 3(b) and (X_1,X_5,X_4) in Fig. 3(c) are three *V*-structures. The BNs shown in Fig. 4 do not have any *V*-structure.

Nakamura et al. [15] established the upper and lower bounds on the dimension $Edim(\mathcal{N})$ of the inner product space for two-valued BNs, where each node is a boolean variable.

Our earlier work in Yang and Wu [25] and Yang and Wu [26] was also focused on these two-valued BNs. For two-valued fully connected Bayesian networks (FBNs) and almost-full Bayesian networks (AFBNs) with n variables, Yang and Wu [25] showed that the VC dimension and the minimum dimension of the inner product space induced by them are $2^n - 1$, and the two dimensional-values induced by a class of two-valued BNs without *V*-structures is $\sum_{i=1}^{n} 2^{m_i} + 1$.

In this paper, we investigate the properties of the inner product space and concept classes induced by multivalued BNs $\mathcal{N}^k$, where each node is a k-valued variable ($X_i \in \{0,1,2,\ldots,k-1\}$). Our work makes two major contributions to the field: (i) for a k-valued FBN with n variables, we show that $VCdim(\mathcal{N}_F^k) = Edim(\mathcal{N}_F^k) = k - 1^n$, and (ii) for k-valued BNs $\mathcal{N}^k$ without *V*-structure, we prove that $VCdim(\mathcal{N}^k) = Edim(\mathcal{N}^k) = (k-1)\sum_{i=1}^{n} k^{m_i} + 1$, where m_i denotes the number of parents for the i^{th} variable. We further derive the upper and lower bounds on the minimum dimension of the inner product space for k-valued non-FBNs.

The results presented in this paper are partially based on our earlier work Yang and Wu [25] and Yang and Wu [26]. For example, some results in Yang and Wu [25] are the special cases of the work in this paper. The rest of the paper is organized as follows. In Section 2, we discuss related work and introduce our notations. In Section 3, we provide the VC dimension and inner product space induced by FBNs. In Section 4, we provide a detailed proof of the upper and lower bounds on the dimension of the inner product space for k-valued BNs. The *VC*-dimension and inner product space induced by BNs without *V*-structure are given in Section 4. We conclude our work in Section 5.

2. Basic concepts and terms

A BN $\mathcal{N}$ consists of a DAG $\mathcal{G} = (V, E)$ and some distributions $\mathcal{P}$ corresponding to $\mathcal{G}$. A topological sort of the nodes (or variables) in a DAG $\mathcal{G} = (V, E)$ is any total ordering of the nodes such that for any pair of nodes X_i and X_j in $\mathcal{G}$, if X_i is an ancestor of X_j, then X_i must precede X_j in the ordering. We assume that every edge $(i,j) \in E$ satisfies $i < j$, that is, E induces a topological ordering on $X_1,X_2,\ldots,X_n$. Given $(i,j) \in E$, X_i is called a parent of X_j and X_j is called a child of X_i. We use PA_i to denote the set of parents of node X_i and let $m_i = |PA_i|$ denote the number of parents. A network $\mathcal{N}$ is fully connected if and only if $PA_i = \{X_1,\ldots,X_{i-1}\}$ holds for every node X_i.

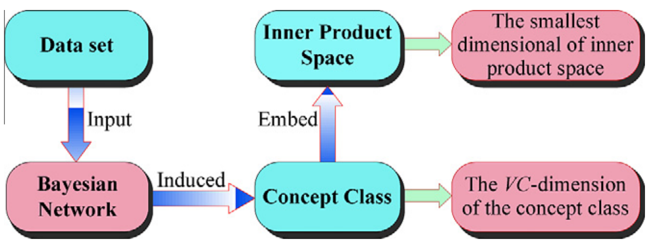

Fig. 1. The relationship is depicted among a BN $\mathcal{N}$, the concept class $\mathcal{C}_{\mathcal{N}}$, Euclidean inner product space and related parameters. By the probability distribution on the BN, the data can be classified. That two dimension-values ($VCdim(\mathcal{N})$ and $Edim(\mathcal{N})$) reflect the classification ability of BN.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات